

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 3, 2001, 11:38:17 ; Search time 16.8 Seconds  
(without alignments)  
459.988 Million cell updates/sec

Title: US-09-441-723-1

Perfect score: 1185

Sequence: 1 MGPLPRTVELFDVLSPSW.....AHLGKWMGPPIPAVNARL 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

- 1: /SIDS1/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1185	100.0	226	21	Novel human glutat
2	1185	100.0	256	20	Human endometrium
3	89	7.5	545	20	Mutant threonine d
4	89	7.5	545	20	Feedback insensiti
5	89	7.5	590	20	Mutant threonine d
6	89	7.5	592	20	Mutant threonine d
7	89	7.5	592	20	Mutant threonine d
8	89	7.5	592	20	Wild type threonin
9	89	7.5	592	20	Arabidopsis wild-t
10	89	7.5	592	20	Feedback insensiti
11	89	7.5	600	20	Mutant threonine d
12	89	7.5	609	20	Mutant threonine d

13	89	7.5	609	20	Y05704
14	83.5	7.0	308	17	W04266
15	83.5	7.0	539	20	Y32942
16	83.5	7.0	539	20	Y05706
17	81	6.8	911	20	Y55957
18	80.5	6.8	339	20	Y33825
19	80.5	6.8	532	20	Y32943
20	80.5	6.8	532	20	Y05707
21	79.5	6.7	1073	18	W32063
22	79.5	6.7	1073	19	W37371
23	78	6.6	341	20	Y37006
24	77	6.5	659	20	Y35014
25	76.5	6.5	216	18	W34203
26	76.5	6.5	216	19	W58004
27	76.5	6.5	317	20	Y39338
28	76.5	6.5	317	20	Y3764
29	76.5	6.5	317	21	Y56814
30	76.5	6.5	497	13	R27786
31	76.5	6.5	497	16	R6693
32	76	6.4	588	20	Y30124
33	75.5	6.4	1398	17	R87008
34	75.5	6.4	1398	18	W24124
35	75.5	6.4	1398	20	W94839
36	74	6.2	342	21	Y91667
37	74	6.2	492	10	P94263
38	74	6.2	554	21	Y91669
39	74	6.2	912	20	Y55939
40	74	6.2	968	20	Y55966
41	73.5	6.2	869	15	R56501
42	73.5	6.2	869	17	W08091
43	73.5	6.2	869	18	W25022
44	73.5	6.2	980	20	W30611
45	73.5	6.2	985	20	W30607

#### ALIGNMENTS

##### RESULT 1

Y77499

ID Y77499 standard; Protein; 226 AA.

AC Y77499;

DT 05-JUN-2000 (first entry)

DE Novel human glutathione S-transferase, GSTS.

XX Glutathione S-transferase; human; GSTS; cancer; immune disorder;  
XX gene therapy; diagnosis; treatment; drug screening.

OS Homo sapiens.

XX US03030809-A.

XX 29-FEB-2000.

XX 25-NOV-1997; 97US-0978174.

XX 25-NOV-1997; 97US-0978174.

XX (INCY-) INCYTE PHARM INC.

XX Hillman JL, Shah P, Lal P, Corley NC;

XX WPI; 2000-205204/18.

XX N-PSDB; Z02599.

XX Isolated nucleic acid encoding glutathione S-transferase useful in the  
XX production of agents for preventing, diagnosing and treating diseases  
XX associated with cell proliferation -

PS Claim 8; Fig 1A-C; 27pp; English.

Feedback insensiti  
Inosine-guanosine  
Mutant threonine d  
Feedback insensiti  
Mouse Sre20-relate  
Amino acid sequenc  
Mutant threonine d  
Feedback insensiti  
Human ST receptor  
Human ST receptor  
Amino acid sequenc  
Chlamydia pneumoni  
Streptomyces non-m  
Streptomyces roseo  
Carboxymethyl cell  
A carboxymethyl ce  
T. maritima thermo  
Adrenodoxin reduct  
Human adrenodoxin-  
A human protein wi  
Protease. Pyrococ  
Pyrococcus furiosu  
W09856926 Seq ID 6  
Human secreted pro  
Bovine adrenal gla  
Human secreted pro  
Human GEK2 protein  
Full length human  
TATA-binding prote  
Human TATA-binding  
TATA-binding prote  
Arabidopsis thalia  
Arabidopsis thalia

XX

CC This sequence represents a novel human glutathione S-transferase, GSTS.  
 CC Nucleotide sequences encoding GSTS were initially isolated from a  
 CC urologic cDNA library, and subsequently extended using cDNA libraries  
 CC derived from other tissues, such as brain or bladder. The present  
 CC sequence is encoded by a consensus cDNA. GSTS, and nucleotides which  
 CC encode it may be used in the prevention, treatment and diagnosis of  
 CC diseases associated with inappropriate GSTS expression, such as cancers  
 CC and immune disorders. Nucleotides which encode GSTS may be used in gene  
 CC therapy to treat disorders associated with reduced expression or activity  
 CC of GSTS, and in antisense therapy for disorders associated with increased  
 CC GSTS expression or activity. They may also be used for the recombinant  
 CC production of GSTS, and as a source of probes and primers to detect and  
 CC quantitate the presence of similar nucleic acid sequences, particularly  
 CC for the diagnosis of GSTS-associated disorders. GSTS proteins may be used  
 CC as antigens in the production of antibodies against GSTS and in assays to  
 CC identify modulators (agonists and antagonists) of GSTS expression and  
 CC activity. The anti-GSTS antibodies and GSTS antagonists may also be used  
 CC to downregulate GSTS expression and activity. Antagonists of GSTS  
 CC expression and function may be used to treat immune disorders (e.g., AIDS,  
 CC anaemia, asthma, Crohn's disease, irritable bowel syndrome, multiple  
 CC sclerosis, osteoarthritis, microbial infections) and cancers (e.g.,  
 CC leukaemia, lymphoma, melanoma, and cancers of the breast, prostate,  
 CC liver, lung and brain). The anti-GSTS antibodies may also be used as  
 CC diagnostic agents.

XX Sequence 226 AA;

Query Match 100.0%; Score 1185; DB 21; Length 226;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-122;  
 Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFYDVLSPYSWLGFEILCRYQNTWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 DB 1 MGPLPRTVELFYDVLSPYSWLGFEILCRYQNTWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 QY 61 RKGLYMANDLKLRLHLLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRE 120  
 DB 61 RKGLYMANDLKLRLHLLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRE 120  
 QY 121 LWMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTERACRYGA 180  
 DB 121 LWMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTERACRYGA 180  
 QY 181 FGLPITVAHVVDGQTHMLFGSDRMELLHLGKWMGPPIPPAVNARL 226  
 DB 181 FGLPITVAHVVDGQTHMLFGSDRMELLHLGKWMGPPIPPAVNARL 226

RESULT 2

Y59988  
 ID Y59988 standard; Protein; 256 AA.

XX AC Y59988;

XX DT 31-JAN-2000 (first entry)

XX DE Human endometrium tumour EST encoded protein 48.

XX KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
 XX treatment; uterine; gene therapy; expressed sequence tag.

XX OS Homo sapiens.

XX PN DE19817948-A1.

XX PD 21-OCT-1999.

XX PF 17-APR-1998; 98DE-1017948.

XX PR 17-APR-1998; 98DE-1017948.

XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX DR N-PSDB; 241996.

XX WPI; 1999-591957/51.

XX New nucleic acid sequences expressed in uterine cancer tissues, and  
 XX derived polypeptides, for treatment of uterine and endometrial cancer  
 XX and identification of therapeutic agents -

XX Claim 23; Page 294; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
 CC that are highly expressed in uterine tumour tissue and which have  
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
 CC are used (i) to identify agents suitable for treatment of uterine or  
 CC endometrial cancer; (ii) directly for treating these forms of cancer  
 CC (including expression from gene therapy vectors) and (iii) for generation  
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
 CC sequence tags) from a particular tissue type before comparison of  
 CC expression patterns. This allows a significantly longer fragment of the  
 CC gene to be revealed, so should reduce the number of failures associated  
 CC with the fact that ESTs from different libraries may represent different  
 CC parts of the same unknown gene, distorting the estimated frequency of  
 CC occurrence in a particular tissue. Y59941-Y60328 represent protein  
 CC fragments encoded by the human endometrium tumour cDNA library derived  
 CC EST fragments represented in Z41981-Z42121.

XX Sequence 256 AA;

Query Match 100.0%; Score 1185; DB 20; Length 256;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-122;  
 Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFYDVLSPYSWLGFEILCRYQNTWNLQRLPSLITGIMKDSGNKPPGLLP 60  
 DB 31 MGPLPRTVELFYDVLSPYSWLGFEILCRYQNTWNLQRLPSLITGIMKDSGNKPPGLLP 90  
 QY 61 RKGLYMANDLKLRLHLLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRE 120  
 DB 91 RKGLYMANDLKLRLHLLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRE 150  
 QY 121 LWMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTERACRYGA 180  
 DB 151 LWMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNLKETTERACRYGA 210  
 QY 181 FGLPITVAHVVDGQTHMLFGSDRMELLHLGKWMGPPIPPAVNARL 226  
 DB 211 FGLPITVAHVVDGQTHMLFGSDRMELLHLGKWMGPPIPPAVNARL 256

RESULT 3

Y32947  
 ID Y32947 standard; Protein; 545 AA.

XX AC Y32947;

XX DT 09-NOV-1999 (first entry)

XX DE Mutant threonine dehydratase/deaminase protein sequence.

XX KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 XX molecular marker; isoleucine toxic structural analog resistance;  
 XX isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 XX polyhydroxybutyrate; antibiotic resistance marker; mutcin.

XX OS Arabidopsis thaliana.

XX OS Synthetic.

XX PN WO9941395-A1.

XX 19-AUG-1999.  
 XX 08-JAN-1999; 99WO-US00560.  
 XX 10-JUL-1998; 98WO-US14362.  
 XX 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
 PI WPI; 1999-527375/44.  
 XX N-PSDB; Z11205.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Claim 13; Page 131-134; 194pp; English.  
 XX This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrates, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX Sequence 545 AA;

Query Match 7.5%; Score 89; DB 20; Length 545;  
 Best Local Similarity 22.8%; Pred. No. 0.3;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDLKLRRHLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 DB 5 ppklplpr-----lkvspnslq-----ypagylgavpntneangsaameyitni 52  
 QY 105 -----AVNLEHP-EMLEKASRELMMRVNSRNEDITE-----PQSTILAA 141  
 DB 53 lsktvydiaiesplqlaklskrigvrmlykredlqpvsfklrgaynmvmkkipadqlak 112  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKETEACRYGAFGLPITVA 188  
 DB 113 gvicssagnhaqgvalasklgctavimpvttpeikwqavnl----- 156  
 QY 189 HVDGQTHMFGSDRMELLALH-----LGEKWMGPPIP 220  
 DB 157 ---gatvvlfgdydqagahakiraeegltfipp 188

RESULT 4  
 Y05711  
 ID Y05711 standard; Protein; 545 AA.

XX Y05711;

XX 19-JUL-1999 (first entry)

XX Feedback insensitive threonine dehydratase/deaminase.

XX Threonine dehydratase/deaminase; OMRI gene; feedback inhibition;  
 KW transgenic plant; selectable marker; isoleucine; mutant.  
 XX

OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 XX Region 439..457  
 FT /note= "regulatory region R4"  
 FT Region 489..507  
 FT /note= "regulatory region R6"  
 FT Misc-difference 452  
 FT /note= "Arg in wild-type TD"  
 FT Misc-difference 497  
 FT /note= "Arg in wild-type TD"  
 XX WO9902656-A1.  
 PN 21-JAN-1999.  
 XX 10-JUL-1998; 98WO-US14362.  
 XX 17-FEB-1998; 98US-0074875.  
 PR 10-JUL-1997; 97US-0052096.  
 XX (PURD ) PURDUE RES FOUND.  
 PA Mourad GS;  
 XX WPI; 1999-120860/10.  
 XX N-PSDB; X25340.

New sequences encode mutant threonine dehydratase/deaminase - which  
 is insensitive to feedback inhibition, useful as a selective marker  
 to produce transformed cells resistant to toxic isoleucine analogues  
 Disclosure; Page 71-73; 120pp; English.

The present sequence represents an Arabidopsis thaliana mutant  
 threonine dehydratase/deaminase (TD) protein which, unlike  
 wild-type TD, is insensitive to feedback inhibition by isoleucine.  
 Claimed polynucleotides (see X25332-40), originally isolated and  
 cloned from A. thaliana mutated line GM1b (omr1/omr1), encode  
 feedback insensitive TD that can be used to transform a wide  
 variety of plants, fungi, bacteria and yeast. Mutant TD differs  
 from the wild-type enzyme only by an R499C amino acid substitution  
 in regulatory region R4, and by an R544H substitution in regulatory  
 region R6. Mutant TD is not only insensitive to feedback  
 inhibition by isoleucine, but is also insensitive to structural  
 analogues of isoleucine that are toxic to plants and microorganisms  
 which synthesize only wild-type TD. Nucleotide sequences encoding  
 mutated forms of TD can therefore be used to create cells that are  
 insensitive to compounds normally toxic to cells expressing only  
 wild-type TD enzymes, and thus may be used to provide a biochemical  
 selectable marker. Transformants that express the mutant TD show  
 increased levels of isoleucine production, and thus provide an  
 improved nutrient source.

Sequence 545 AA;

Query Match 7.5%; Score 89; DB 20; Length 545;  
 Best Local Similarity 22.8%; Pred. No. 0.3;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRRHLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 DB 5 ppklplpr-----lkvspnslq-----ypagylgavpntneangsaameyitni 52  
 QY 105 -----AVNLEHP-EMLEKASRELMMRVNSRNEDITE-----PQSTILAA 141  
 DB 53 lsktvydiaiesplqlaklskrigvrmlykredlqpvsfklrgaynmvmkkipadqlak 112  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKETEACRYGAFGLPITVA 188  
 DB 113 gvicssagnhaqgvalasklgctavimpvttpeikwqavnl----- 156

QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
 ID Y32950 standard; Protein; 590 AA.  
 AC Y32950;  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE  
 KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX  
 PN WO9941395-A1.  
 XX  
 PD 19-AUG-1999.  
 XX  
 XX 08-JAN-1999; 99WO-US00560.  
 XX  
 PF 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 PA  
 PI Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
 XX  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11208.  
 XX  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX  
 XX Example 3; Page 119-123; 19app; English.

XX This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX  
 XX Sequence 590 AA;

Query Match 7.5%; Score 89; DB 20; Length 590;  
 Best Local Similarity 22.8%; Pred. No. 0.34;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 Db 50 ppklplpr-----lkvspnslq---ypagylgavpntneagsiaameyltni 97  
 QY 105 -----AVNLEHP-EMLEKASRELWMRWRSNEDITE-----PQSILAA 141

Db 98 lskvydiaiesplqlaklskrlgvrmylkrldlqpvsfklrgaynmvmvklpadqlak 157  
 QY 142 AEKAGMSAQOAG-----LLEKIATPKVKNLKETTEACRYCAGLPINVA 188  
 Db 158 gvicssagnhaqgvalsasklgtctavimvpttpeikwqavenl----- 201  
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
 Db 202 ---gatvvifgdydqaqahakiraeegltfipp 233

## RESULT 6

Y32939  
 ID Y32939 standard; Protein; 592 AA.  
 XX  
 AC Y32939;  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE  
 KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX  
 PN WO9941395-A1.  
 XX  
 PD 19-AUG-1999.  
 XX  
 PF 08-JAN-1999; 99WO-US00560.  
 XX  
 PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 PA  
 PI Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
 XX  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11197.  
 XX  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX  
 XX Claim 13; Fig 8; 19app; English.

XX This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.  
 XX  
 XX Sequence 592 AA;

Query Match 7.5%; Score 89; DB 20; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 0.34;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

```

SQ      Sequence          592 AA;

Query Match              7.5%; Score 89; DB 20; Length 592;
Best Local Similarity    22.8%; Pred. No. 0.34;
Matches                  49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY   55 PPGL-LPRKGLYMANDLKLRHHHQIPIHPKPDELNV-----LEKGSLS-AMRELT-- 104
     ||| |||         ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: |
Db   52 ppklpplr-----lksvpsnlq-----ypagylgavpertneagslaeameyitnl 99

QY   105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----PQSILAA 141
     : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   100 lskvydiaiesplglakklslkrigvrmylkrediqpvfsfklrgaynmvmklpadqlak 159

QY   142 AEKAGSAEQAQG-----LLEKIATPKVKKNQLKETTEAACRYGAFGLPTVA 188
     . : ||| : : : : : : : : : : : : : : : : : : : : : :
Db   160 gvicsagnhagvalsasklgctavimpyttpeikwqavenl----- 203

QY   189 HVDGQTHLMFGSDRMELLAHL-----LGKKWMGPPIP 220
     | : ||| : : : : : : : : : : : : : : : : : : : : : :
Db   204 --gatvvlfgsydqaqhahakraeeegitflpp 235

RESULT 8
Y32951 ID Y32951 standard; Protein; 592 AA.
XX AC Y32951;
XX XX
XX DT
XX DE Wild type threonine dehydratase/deaminase protein sequence.
KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;
KW molecular marker; isoleucine toxic structural analog resistance;
KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
KW polyhydroxybutyrate; antibiotic resistance marker.
XX OS Arabidopsis thaliana.
XX PN WO9941395-A1.
XX PD 19-AUG-1999.
XX PF 08-JAN-1999; 99WO-US000560.
XX PR 10-JUL-1998; 98WO-US14362.
XX PR 17-FEB-1998; 98US-0074875.
XX PA (DOWC ) DOW AGROSCIENCES LLC.
XX PA (PURD ) PURDUE RES FOUND.
XX PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;
XX WPI; 1999-527375/44.
XX N-PSDB; Z11209.
XX PT New nucleic acid encoding threonine dehydratase/deaminase resistant
XX to feedback inhibition, useful as selection marker for cell
XX transformation and to impart herbicide resistance
XX Disclosure; Page 86-89; 194pp; English.
CC This sequence is the wild type Arabidopsis thaliana threonine
CC dehydratase/deaminase (TD) protein. The invention relates to mutants of
CC the encoded protein, that are feedback insensitive TD mutants. The TD DNA
CC sequence is used as molecular marker (impairing resistance to toxic
CC structural analogues of isoleucine) for selecting transformed cells and
CC to produce transformants with increased levels of isoleucine (and thus
CC better nutritional value) or of intermediates in biosynthesis of
CC isoleucine (e.g. 2-oxobutyrate), for synthesis of the degradable
CC biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use

```







CC structural analogues of isoleucine that are toxic to plants and  
 CC microorganisms which synthesize only wild-type TD. Nucleotide  
 CC sequences encoding mutated forms of TD can therefore be used to  
 CC create cells that are insensitive to compounds normally toxic to  
 CC cells expressing only wild-type TD enzymes, and thus may be used  
 CC to provide a biochemical selectable marker. Transformants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX  
 SQ Sequence 609 AA;

Query Match 7.5%; Score 89; DB 20; Length 609;  
 Best Local Similarity 22.8%; Pred. No. 0.36;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 Db 69 ppkpplr-----lkvspnslq-----ypagylgavperrneangsiaameyltni 116  
 QY 105 -----AVNLEHP-EMLEKASRELMMRVMSRNEDEITE-----PQSILAA 141  
 Db 117 lstkvydiaiesplqlaklsrlgrvmylkredlqpvfsklrgaynmvklpadqlak 176  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVNQLKETTEAACYGAFGLPITVA 188  
 Db 177 gvlcssagnhaggvalsasklgctavimpvttpeikwqavenl----- 220  
 QY 189 HVDGOTMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 Db 221 ---gatvvlfdsydaqahakiraeegltfipp 252

RESULT 14  
 W04266  
 ID W04266 standard; Protein; 308 AA.  
 AC W04266;  
 XX  
 DT 30-JUN-1997 (first entry)  
 DE  
 DE Inosine-guanosine kinase.  
 XX  
 KW Inosine-guanosine kinase; 5'-inosinic acid; 5'-guanylic acid;  
 KW Corynebacterium ammoniagenes; ATP; seasoning; food.  
 XX  
 OS Exiguobacterium sp. (ATCC 35652).  
 XX  
 PN W09630501-A1.  
 XX  
 PD 03-OCT-1996.  
 XX  
 PF 22-MAR-1996; 96WO-JP00761.  
 XX  
 PR 09-JUN-1995; 95JP-0177900.  
 PR 24-MAR-1995; 95JP-0102888.  
 XX  
 PA (AJIN ) AJINOMOTO CO INC.  
 XX  
 PI Kawasaki H, Shimaoka M, Usuda Y, Utogawa T;  
 XX  
 DR WPI; 1996-455349/45.  
 DR N-PSDB; T33972.  
 XX  
 PT Prodn. of 5'-inosinic acid or 5'-guanylic acid - using C.  
 PT ammoniagenes transformed by inosine-guanosine kinase gene, for use  
 PT in food seasoning  
 XX  
 PS Example 14; Page 65-66; 72pp; Japanese.  
 XX  
 CC This sequence represents the inosine-guanosine kinase sequence derived  
 CC from Exiguobacterium sp. This sequence was used in the production of

CC 5'-inosinic acid or 5'-guanylic acid by reacting an organism, pref.  
 CC Corynebacterium ammoniagenes, capable of regenerating ATP, containing  
 CC this recombinant DNA, with inosine, guanosine or their precursors, and  
 CC an energy and phosphoric acid source. 5'-inosinic acid and 5'-guanylic  
 CC acid are useful as seasonings for food.  
 XX  
 SQ Sequence 308 AA;

Query Match 7.0%; Score 83.5; DB 17; Length 308;  
 Best Local Similarity 25.8%; Pred. No. 0.52;  
 Matches 40; Conservative 25; Mismatches 57; Indels 33; Gaps 8;  
 QY 63 GLYMANDLKLRLHLLQIPHPKDFLSVMLEKGSLSAMRFLTAVNLEHP-EMLEKASRELW 122  
 Db 40 grnvaqlgvlndvrfvstvtndqgi-----gvleelrsln-vvnehvdlledngmgw 94  
 QY 123 MRVMSRNEDEITEPQSILAAAEKAGMSAEQAQGLEKLEKIATPKVK-----NOLKE 170  
 Db 95 lavmdnngdl--qtsiskpdeamme---qcilrridvtfaestavaidldslsvnlne 148  
 QY 171 TTEAACR-----YGAFGLPITVAHVDCGQTHMLFG 199  
 Db 149 tie-lcremkplpygvvgc---hlsviernrhliqg 179

RESULT 15  
 Y32942  
 ID Y32942 standard; Protein; 539 AA.  
 XX  
 AC Y32942;  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 XX  
 KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polynhydroxybutyrate; antibiotic resistance marker; muteln.  
 XX  
 OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX  
 PN W09941395-A1.  
 XX  
 PD 19-AUG-1999.  
 XX  
 PF 08-JAN-1999; 99WO-US00560.  
 XX  
 PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX  
 PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX  
 DR WPI; 1999-527375/44.  
 DR N-PSDB; Z11200.

New nucleic acid encoding threonine dehydratase/deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX  
 PS Claim 13; Page 106-109; 194pp; English.  
 XX  
 CC This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of

Sequence 539 AA;

	QY	59	LPRKGLYMANDKLRLHHHLQIPHFHPKDFLSVM-----LEKGSLS-AMRELT-----	104
			:    :	
			:     :     :     :	
	Db	4	lpr -----lkvspnslq---ypagylgapertneaengsiaameylnilstkv	51
			:	
	QY	105	-AVNLEHP-EMLEKASRELWMRWVSREDITE-----PQTSLAAAEAKG	146
			:     :     :     :     :	
	Db	52	ydiaesplqlakklskrlgvrmylkredlpvfysfklrgaynmvklpadqlakgvics	111
			:	
	QY	147	MSAEQAOA-----LLEKIATPKVKQLKETTEAACRYCAGFGLPTVAHVDSQ	193
			:     :     :     :     :	
	Db	112	sagnhaqqavalsasklgctavimpyvttpaikwgavenl-----ga	152
			:	
	QY	194	THMLFGSDRMELLAHL----LGSKWGMPIPP	220
			:     :     :	
	Db	153	tvwifqdsyddqaahakahakraeseqltfipp	182

Search completed: April 3, 2001, 11:39:01  
Job time: 44 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 3, 2001, 11:38:17 ; Search time 14.7 Seconds  
(without alignments)  
276.074 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPLRTRVELFDVLSFYSW.....AHLIGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1185	100.0	226	3	US-08-978-174-1
2	874	73.8	226	3	US-08-978-174-3
3	80.5	6.8	339	2	US-08-855-714-3
4	76.5	6.5	317	2	US-09-066-075-2
5	76.5	6.5	317	2	US-08-518-615A-2
6	76.5	6.5	317	3	US-08-951-889-2
7	76.5	6.5	497	1	US-08-075-193-4
8	76.5	6.5	497	2	US-08-564-090A-4
9	76.5	6.5	497	4	PCT-US94-06698-4
10	75.5	6.4	1398	1	US-08-750-532-9
11	73.5	6.2	523	2	US-08-473-553A-3
12	73.5	6.2	869	1	US-08-188-582-32
13	73.5	6.2	869	1	US-08-646-715-32
14	73.5	6.2	980	2	US-08-473-553A-6
15	73.5	6.2	985	2	US-08-473-553A-2
16	73	6.2	948	1	US-08-698-551-14
17	73	6.2	948	2	US-08-602-228-14
18	73	6.2	948	2	US-08-533-901B-14
19	73	6.2	948	2	US-08-839-032A-14
20	73	6.2	948	4	PCT-US95-12724-14
21	71.5	6.0	543	2	US-08-922-170B-10
22	71.5	6.0	587	1	US-07-955-905A-23
23	70.5	5.9	615	2	US-08-484-101B-38
24	70.5	5.9	401	1	US-08-198-446B-11
25	70	5.9	401	2	US-08-870-693-11
26	69.5	5.9	259	2	US-07-857-224B-51
27	69.5	5.9	529	4	PCT-US95-05008-15
28	69.5	5.9	529	4	PCT-US95-05008-15

Sequence 2, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 10, Appli  
Sequence 50, Appli  
Sequence 14, Appli  
Sequence 4, Appli  
Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-978-174-1  
; Sequence 1, Application US/08978174  
; Patent No. 6030809  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Purvi  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,174  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0430 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 226 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADTUT04  
; CLONE: 1554593  
; US-08-978-174-1

Query Match

100.0%; Score 1185; DB 3; Length 226;



Db 47 PLPTSPMNLKMCOSAPPTSLKRLRFLCPRPHWK-----SLRTASQKSSFTKAALK 99  
QY 60 PRK-GLYMANDLKLRLHQLPIHPKDFLSVMLEKGS-----SAMRFLTAVNLEH 110  
Db 100 PRGGSWMTKRSPRALGAWH-PQRRRLGYDGKGQVRLASLDQACNAFAAIN-KA 157  
QY 111 PEMLE---KASRELWMRVWSRNEIDITEPOSILAAAEKAG-----146  
Db 158 PALLEGFVEFEREV-----SVIAARDRSGNVAIFDLAENVHKDGLATST 202  
QY 147 -----MSAEQAQGLLEKIA-----TPKVNQLKETTE 173  
Db 203 VPAAISVQTAAEAARTAAEKLLHALDYGVGLGLEFVLKDGTLTLLANEFAPRVHNS-GHWTE 261  
QY 174 AACRYGAF-----GLPI--TVAHVDGQTHMLFGSD-----RMELLAHLLEK 213  
Db 262 AACAIQFQEHIRAVAGLPLGNTDRHSDCVMENLIGDDIEKVPAILCERNAVLHLYGKK 320

RESULT 4  
US-09-066-075-2  
; Sequence 2, Application US/09066075  
; Patent No. 5925749  
; GENERAL INFORMATION:  
; APPLICANT: Mathur, E., et al.  
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,075  
; FILING DATE:  
; CLASSIFICATION:  
; APPLICATION NUMBER: US/08/518,615  
; FILING DATE: August 23, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 331400-20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 317 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-09-066-075-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
QY 76 HLQIPIH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEK 116  
Db 49 HVRIPIRWSTHAYAFPYPKIMDRFKRVDEVINGALKRG-----LAVAINIHVEELMN 102  
QY 117 ASRE-----LWMRVWSRNE-----ITEPOS-----137

Db 103 DPEEHKRFRLWKQIADRYKDYPETLFFELNPEHGNLTPEKWNELLEALKVIRSIDK 162  
QY 138 -----ILAAAEKAGSAEQAGLLEKIATPKVNQLKETTEAACRYGAFGLPITVAHVVDGQ 193  
Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE---RNSIVTIHYINPFEF-----203  
QY 194 THMLFGSDRMELLALHLLGEEKWGP 217  
Db 204 THQ--GAEWEGSEKWLGRKWGSP 225

RESULT 5  
US-08-518-615A-2  
; Sequence 2, Application US/08518615A  
; Patent No. 5962258  
; GENERAL INFORMATION:  
; APPLICANT: Mathur, E., et al.  
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/518,615A  
; FILING DATE: August 23, 1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 331400-20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 317 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-518-615A-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
QY 76 HLQIPIH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEK 116  
Db 49 HVRIPIRWSTHAYAFPYPKIMDRFKRVDEVINGALKRG-----LAVAINIHVEELMN 102  
QY 117 ASRE-----LWMRVWSRNE-----ITEPOS-----137  
Db 103 DPEEHKRFRLWKQIADRYKDYPETLFFELNPEHGNLTPEKWNELLEALKVIRSIDK 162  
QY 138 -----ILAAAEKAGSAEQAGLLEKIATPKVNQLKETTEAACRYGAFGLPITVAHVVDGQ 193  
Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE---RNSIVTIHYINPFEF-----203



STREET: 3000 EL CAMINO REAL  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/564.090A  
FILING DATE: 02/05/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD L. NEELEY, PH.D.  
REGISTRATION NUMBER: 30,092  
REFERENCE/DOCKET NUMBER: UCAL-236/01US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-843-5000  
TELEFAX: 415-857-0663  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-564-090A-4

Query Match 6.5%; Score 76.5; DB 2; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;

QY 14 VLSPSYSLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPGLLPRKG-LYMANDLKL 72  
Db 195 LITPPEHLEALLLCQRTDITKAAL-----GVLRQSRVKTVMVLRGRRGPLOVAFTIKE 246  
QY 73 LRHLQI-----PIHPPKDFL-----SYMLEKGSLSAMRFLTAVNLEHPMELEKASREL 121  
Db 247 LREMIQPGARPILDVDFGLQDKIKEVPRPRKRLTELLRTAT--EKPGPAEAAARQAS 304  
QY 122 WMRVWSRNEIDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKYKNOLKETEEA 174  
Db 305 ASRAWGL-RFRSPQQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 9  
PCT-US94-06698-4  
Sequence 4, Application PC/TUS9406698  
GENERAL INFORMATION:  
APPLICANT: MILLER, WALTER L.  
APPLICANT: HARIKRISHNA, JENNIFER A.  
APPLICANT: BLACK, STEPHEN M.  
TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROBBINS, BERLINER & CARSON  
STREET: 201 NORTH FIGUEROA STREET  
CITY: LOS ANGELES  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 90012  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/06698  
FILING DATE: FILED HERewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: BERLINER, ROBERT  
REGISTRATION NUMBER: 20,121  
REFERENCE/DOCKET NUMBER: 5555-224-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213-977-1001  
TELEFAX: 213-977-1003  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-06698-4

Query Match 6.5%; Score 76.5; DB 4; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;

QY 14 VLSPSYSLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPGLLPRKG-LYMANDLKL 72  
Db 195 LITPPEHLEALLLCQRTDITKAAL-----GVLRQSRVKTVMVLRGRRGPLOVAFTIKE 246  
QY 73 LRHLQI-----PIHPPKDFL-----SYMLEKGSLSAMRFLTAVNLEHPMELEKASREL 121  
Db 247 LREMIQPGARPILDVDFGLQDKIKEVPRPRKRLTELLRTAT--EKPGPAEAAARQAS 304  
QY 122 WMRVWSRNEIDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKYKNOLKETEEA 174  
Db 305 ASRAWGL-RFRSPQQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 10  
US-08-750-532-9  
Sequence 9, Application US/08750532  
Patent No. 5756339  
GENERAL INFORMATION:  
APPLICANT: MITTA, Masanori  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MORISHITA, Mio  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
STREET: 419 Seventh Street N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,532  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01095  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/130236  
FILING DATE: 13-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/173912  
FILING DATE: 26-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618



; MOLECULE TYPE: protein  
US-08-188-582-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. NO. 15;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;  
QY 32 IWNLNQLRPSLITGIMKDSGNKPPGLPRKGLYMANDLKLRLHHLQIPIHFHFKDFLSVM 91  
DB 318 VWFKGKOWPTLLOAMQVEKATGTSLSLSP-----HLPGELAICRSRGAVCLWSPEDGLRQI 373  
QY 92 LEKGSLSAMRELTAVNLEHPEMLEKASRELMMRWVSRNEDITEPQSIILAAAEKAGMSAEQ 151  
DB 374 -----YRDPETLVPRDSSWR--WA---DFTAHPRVLTGVDRGTGVMKLD 412  
QY 152 AQG-----LLEKATPKVKNLKETTAAACRYGA-----FG-----LPITVAHVVDG 192  
DB 413 TQGGPPGCGLLFLRLG-----AEASCKQGERVLLTQYLGHSSPKCLPPTLHLVCT 461  
QY 193 QTHMFGSDRMELLHLGELKWMGPIP-PAVNARL 226  
DB 462 QFSLYLVDERLPLVPM-----KWNHGLPSPLLIARL 493

RESULT 13  
US-08-646-715-32  
; Sequence 32, Application US/08646715  
; Patent No. 5637686  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,715  
; FILING DATE: 09-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/188,582  
; FILING DATE: 28-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 869 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-646-715-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. NO. 15;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;  
QY 32 IWNLNQLRPSLITGIMKDSGNKPPGLPRKGLYMANDLKLRLHHLQIPIHFHFKDFLSVM 91  
DB 318 VWFKGKOWPTLLOAMQVEKATGTSLSLSP-----HLPGELAICRSRGAVCLWSPEDGLRQI 373  
QY 92 LEKGSLSAMRELTAVNLEHPEMLEKASRELMMRWVSRNEDITEPQSIILAAAEKAGMSAEQ 151  
DB 374 -----YRDPETLVPRDSSWR--WA---DFTAHPRVLTGVDRGTGVMKLD 412  
QY 152 AQG-----LLEKATPKVKNLKETTAAACRYGA-----FG-----LPITVAHVVDG 192  
DB 413 TQGGPPGCGLLFLRLG-----AEASCKQGERVLLTQYLGHSSPKCLPPTLHLVCT 461  
QY 193 QTHMFGSDRMELLHLGELKWMGPIP-PAVNARL 226  
DB 462 QFSLYLVDERLPLVPM-----KWNHGLPSPLLIARL 493

RESULT 14  
US-08-473-553A-6  
; Sequence 6, Application US/08473553A  
; Patent No. 5859338  
; GENERAL INFORMATION:  
; APPLICANT: Meyerowitz, Elliot M.  
; APPLICANT: Clark, Steven E.  
; APPLICANT: Williams, Robert W.  
; TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
; TITLE OF INVENTION: Transformed Plants, and Proteins  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/473,553A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-60886/RET/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 980 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-473-553A-6

Query Match 6.2%; Score 73.5; DB 2; Length 980;  
Best Local Similarity 21.8%; Pred. NO. 18;  
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 14;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 3, 2001, 11:38:20 : Search time 15.64 seconds  
(without alignments)  
981.173 Million cell updates/sec

Title: US-09-441-723-1

Perfect score: 1185

Sequence: 1 MGPLPRTVEFYDVLSPYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_66: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328.5	27.7	226	2 T27747	hypothetical prote
2	298.5	25.2	225	2 T34201	hypothetical prote
3	198	16.7	195	2 G83629	hypothetical prote
4	164.5	13.9	194	2 S72164	2-hydroxychromene-
5	140	11.8	197	2 T31286	2-nitrotoluene dio
6	125	10.5	33	2 S17164	glutathione trans
7	95	8.0	203	2 C55552	2-hydroxychromene-
8	91	7.7	581	2 H72425	ABC transporter, A
9	90.5	7.6	410	2 C59127	flavoprotein A hom
10	89.5	7.6	962	1 SNECPI	pitriylisin (EC 3.4
11	89	7.5	199	2 T49343	probable isomerase
12	89	7.5	592	2 T51712	threonine dehydrat
13	88.5	7.5	613	2 A35296	secretogranin II p
14	87	7.3	926	2 E83375	probable glycosyl
15	84.5	7.1	316	2 B71301	probable tRNA delt
16	84.5	7.1	971	2 S34595	probable membrane
17	83.5	7.0	488	2 A40367	catalase (EC 1.11.
18	83.5	7.0	700	1 S09748	kinesin-related pr
19	82.5	7.0	263	2 T51169	hypothetical prote
20	81.5	6.9	619	2 S02180	secretogranin II -
21	81.5	6.9	1061	1 DJAD12	DNA-directed DNA p
22	81	6.8	274	2 S75896	hypothetical prote
23	80.5	6.8	705	2 A48144	protein kinase CDC
24	80	6.8	402	2 E69107	hypothetical prote
25	80	6.8	540	2 A70358	topoisomerase I -
26	79.5	6.7	547	2 E69647	catalase (EC 1.11.
27	79.5	6.7	1073	1 OXHUHX	heat-stable entero
28	79	6.7	230	2 C75301	rRNA methylase Spo
29	78.5	6.6	333	2 F83215	conserved hypothet

30 78.5 6.6 383 2 G71648  
31 78.5 6.6 493 2 T06060  
32 78 6.6 248 2 G69352  
33 78 6.6 313 2 G72784  
34 78 6.6 338 2 A71545  
35 78 6.6 910 2 C69069  
36 78 6.6 1162 2 T40817  
37 78 6.6 1479 2 T17401  
38 77.5 6.5 482 2 T44628  
39 77.5 6.5 634 2 C64454  
40 77 6.5 253 2 G75008  
41 77 6.5 659 2 D72083  
42 76.5 6.5 184 2 T40215  
43 76.5 6.5 497 1 A40487  
44 76.5 6.5 498 1 JT0751  
45 76.5 6.5 843 2 H82362

alanine dehydrogen  
cellulase (EC 3.2.  
branched-chain ami  
hypothetical prote  
probable o-sialogl  
cation-transportin  
botulinum toxin no  
transcription regu  
probable transposa  
hypothetical prote  
hypothetical prote  
transport ATP bind  
hypothetical prote  
ferredoxin--NADP+  
ferredoxin--NADP+  
adenylate cyclase

#### ALIGNMENTS

RESULT 1

T27747

hypothetical protein ZK1320.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27747

R:Berk, M.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z20414

A:Accession: T27747

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-226 <WIL>

A:Cross-references: EMBL:Z46934; PIDN:CAA87039.1; GSPDB:GN00020; CESP:ZK1320.1

A:Experimental source: clone ZK1320

C:Genetics:

A:Gene: CESP:ZK1320.1

A:Map position: 2

A:Introns: 23/3; 177/3

Query Match

Best Local Similarity 27.7%; Score 328.5; DB 2; Length 226;

Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;

QY 1 MGPLPRTVEFYDVLSPYSWLGFEILCRQNTW-NINLQLRPSLITGIMKDSGNKPPGILL 59

Db 1 MPKLPR-IDFYDVISPSYIAFEVFKLETQWKGVITIRYIPFLGAVMKESGNRPPAML 59

QY 60 PRKGLYMANDLKLRRHHLQIPIHFPRKDFLSVLMLEKSLSAMRFLTAVNLEHPEMLEKASR 119

Db 60 PARSIMMTDLKRYAKFWDIPLTPPLFWEWIKKYRTTGAMKVLVLVLEQDKELMLRAAR 119

QY 120 ELNMRVSRNEDITEPQSILAAAEKAGSAEQGLEKATPKVKQLKETTETACRYG 179

Db 120 EMWVRLWSRSEKIEFDQDFEVLKAVGV--KNPEQIVKSKDEKRYIKILMENTNKGVDLM 177

QY 180 AFGCLP-ITVAHVGDGTHMLFGSDRMELLAHLLEKGMGP 219

Db 178 AYGAPWINVHTEDGSEHSFFGSDRPHLIADLLOQP--QPLP 216

RESULT 2

T34201

hypothetical protein D2024.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T34201

R:Du, Z.; Gattung, S.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid D2024.

A:Reference number: Z21488

```

QY      180  AFGLPITVAHVDGQTHMLFGSDRMELLAHL 210
          |||      : || ||: : |
Db      167  VEGAPSFV---GDQLEFGQDRLDFAEVL 193

```

**QY**      4 L P R T V E Y D V L S P Y S W I G F E I L C R Q N I W N I N L Q R P S L T I G M K D S G N K P P G --- L L P    60  
               :  
**Dd**      1 M T R T I D F Y D F I S P S Y L A Q L K L P E L A R A A G C T V D Y W P I D I P E A K I A A G N Y G P S N R E V L P    60  
  
**QY**      61 R K G L Y M A N D L K L L R H H L O I P I H P K D F ----- L S V M L E K G S L S A M R E L T A V N L E H P E M L    114

Db 61 -KIKVMKADLERWAERYGVPLTFPPASFACADWNCVLFAREHGKAEPVT----- 109  
QY 115 EKASRELWVRVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETEA 174  
Db 110 -----DAYRIWGGIDPGDRNELAACAIAGL---DPAALIAFVESPAGNEYKRSQ 161  
QY 175 ACRYGAFGLPITVAHVDTGTHMLFGSDRMELLALHLL 210  
Db 162 AIQGVYCAPL--MFVDDQ--IFWGNDRDLFLAEYL 193

RESULT 6  
S17164  
glutathione transferase (EC 2.5.1.18) 13 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Nov-1993 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
C:Accession: S17164  
R:Harris, J.M.; Meyer, D.J.; Coles, B.; Ketterer, B.  
Biochem. J. 278, 137-141, 1991  
A:Title: A novel glutathione transferase (13-13) isolated from the matrix of rat liver  
A:Reference number: S17164; MUID:91354194  
A:Accession: S17164  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-33 <HAR>  
C:Keywords: transferase

Query Match 10.5%; Score 125; DB 2; Length 33;  
Best Local Similarity 71.9%; Pred. No. 0.00013;  
Matches 23; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 PLPRVELFYDVLSPYSWLGFEILCRYQNIWN 34  
Db 2 PAPRVLEFYDVLSPYSXGLFEVLXRYQHLXN 33

RESULT 7  
C5552  
2-hydroxychromene-2-carboxylate isomerase - Pseudomonas putida plasmid NAH7  
C:Species: Pseudomonas putida  
C:Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Sep-1999  
C:Accession: C5552  
R:Eaton, R.W.  
J. Bacteriol. 176, 7757-7762, 1994  
A:Title: Organization and evolution of naphthalene catabolic pathways: sequence of the  
from the NAH7 plasmid  
A:Reference number: A5552; MUID:95095951  
A:Accession: C5552  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-203 <EAT>  
A:Cross-references: GB:U09057; NID:g483790; PIDN:AAA66358.1; PID:g483793  
C:Genetics:  
A:Gene: nahD  
A:Genome: plasmid  
A:Start codon: GTG  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 8.0%; Score 95; DB 2; Length 203;  
Best Local Similarity 24.2%; Pred. No. 0.53;  
Matches 51; Conservative 35; Mismatches 97; Indels 28; Gaps 8;

QY 8 VELFYDVLSPYSWLGFEILCRYQNIWNINLQRLPSLITGIMKDSGNKPPCL--LPRKGLY 65  
Db 3 VDFYDFLSPFSYLIANQSLAQDGLTIRYNADLARVKTIAIGNVGSNRDLKVKLDY 62  
QY 66 MANDLKLLRHHLQPIHFPPKDFLSVMLEKG---SLSAMRFLTAVNLEHPMLEKASRELW 122  
Db 63 LKYDLQRLWAQILGIPLFPANYSRRMNI GFYYSAGAEAAAYVNV-----VF 110

QY 123 MRVMSRNEDETEPOSILAA--AEKAGMSAEQAQGLLEKIATPKVKNQKLTETEAACRYGA 180  
Db 111 NAVW--GEGIAFDLESPLALVSEKLGWDRSAFHEFLSSNAATE---RYDEQTHAAIERKV 165  
QY 181 FGLPITVAHVDTGTHMLFGSDRMELLALHLLG 211  
Db 166 FGVP-TMFLGD---EMWGNDRFLFLESAMG 192

RESULT 8  
H72425  
ABC transporter, ATP-binding protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72425  
R:Nelson, M.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.;  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from geno  
A:Reference number: A72200; MUID:99287316  
A:Accession: H72425  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-581 <ARN>  
A:Cross-references: GB:AE001691; GB:AE000512; NID:g4980517; PIDN:RAD35137.1; PID:  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0043  
C:Superfamily: Escherichia coli ABC transporter mdIA; ATP-binding cassette homolog

Query Match 7.7%; Score 91; DB 2; Length 581;  
Best Local Similarity 19.0%; Pred. No. 4.5;  
Matches 35; Conservative 44; Mismatches 67; Indels 38; Gaps 6;

QY 28 RYQNIW-----NINQLRPSLITGIMKDSGNKPPGL-----LPRKGLYMAN 68  
Db 343 RFENVWFSYDGKNWLVKDINDLPQFKLYAIVGTGGGKSTLMSLINGLYIPQGNFID 402  
QY 69 DLKLLRHHLQIP---IHFPRDEL-----SVMLEKGSLSAMRFLTAVNLEHP-EMLE 115  
Db 403 EPLLEYNLKLVRKQIAAVQDVLVLFSCITLDNIRLFDSEIPERVLEALKRVHALDIE 462  
QY 116 KASRELWVRVSRNEDITEPOSILAAAEKAGM-----SAEQOGLLEKIATPKVKNQK 169  
Db 463 RLPGGVYVEIVERGTTLSAGERQLIALARAVLFDKAFILDEATSNVDVITETKIOBALE 522  
QY 170 ETE 173  
Db 523 ELSK 526

RESULT 9  
C69127  
flavoprotein A homolog (II) - Methanobacterium thermoautotrophicum (strain Delta  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: C69127  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredg  
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiw  
K.L.S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H  
A:Reference number: A69000; MUID:98037514  
A:Accession: C69127  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-410 <MTH>  
A:Cross-references: GB:AE000809; GB:AE000666; NID:g2621265; PIDN:AAB84726.1; PID:  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH220

A:Start codon: TTG  
C:Superfamily: Methanobacterium flavoprotein A  
C:Keywords: flavoprotein

Query Match 7.6%; Score 90.5; DB 2; Length 410;  
Best Local Similarity 25.1%; Pred. No. 3.2;  
Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;  
QY 13 DVLSPYSLWGFELC-----RQNTWNLQLRPLSLITGIMKDSGNK-----54  
Db EVLKRI---GSELICTAKAAGLRQHYSPDTPMQ---TVKTDGSDTDLGGKTLTLEAP 152  
QY 55 ----PPG---LPRKGLYMANDLKLRLHQLIPHFDPKDLFSLVMLEKGSLSAMRF---L 103  
Db 153 MLHWPDSMFTLLEEGILFSND--AQGHLCISKRFDKDPVAVLMD---AAKFFVANLL 207  
QY 104 TAVNLEHPEMLEKAS---RELWM-----RVMSRNEDITEPQSILAAAEKAGMSA 149  
Db 208 TPLS---PLVLRKFSEVKELGLEKIGMIAPSHQIWI-----TEPLKIIAA-----YT 252  
QY 150 EQAGLLEKIAIPKVNQLKTEEAACRYGAGLPITVAHVGDQTHMLFGSDRMELLALH 209  
Db 253 DWATGCKRDKAT-IYDTWHYSTRMLAHMAEGL--MAADVDSMHFLHEDERSEIVKNI 309  
QY 210 LGKE-----WMGPIP 219  
Db 310 LESKAVFIGSPTMFNGPFP 328

RESULT 10  
SNECPI  
Pitriylsin (EC 3.4.24.55) precursor [validated] - Escherichia coli  
N:Alternate names: endopeptidase Pi; proteinase III  
C:Species: Escherichia coli  
C>Date: 31-Mar-1993 #sequence\_revision 31-Oct-1997 #text\_change 18-Feb-2000  
C:Accession: F65064; A29093; A25765; B25532  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: F65064  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-962 <BLAT>  
A:Cross-references: GB:AE000365; GB:U00096; NID:q2367163; PIDN:AA24436.1; PID:g2367164;  
A:Experimental source: strain K-12, substrain MG1655  
R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.  
Gene 54, 185-195, 1987  
A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Escherichia coli K-12.  
A:Reference number: A29093; MUID:88005781  
A:Accession: A29093  
A:Molecule type: DNA  
A:Residues: 1-276, 'HYHSLR', 283, 'W', 285-296 <CLA>  
A:Cross-references: GB:M17095; NID:g147390; PIDN:AAA24436.1; PID:g147391  
A:Experimental source: strain K12  
A:Note: Part of this sequence, including the amino end of the mature protein, was confirmed by Nucleic Acids Res. 14, 7695-7703, 1986  
A:Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding protease III.  
A:Reference number: A25765; MUID:87040734  
A:Accession: A25765  
A:Molecule type: DNA  
A:Residues: 1-962 <FIN>  
A:Cross-references: GB:X06227; NID:g42560; PIDN:CAA29576.1; PID:g42561  
R:Becker, A.B.; Roth, R.A.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992  
A:Title: An unusual active site identified in a family of zinc metalloendopeptidases.  
A:Reference number: A38854; MUID:92237263  
A:Contents: annotation; active site  
C:Genetics:  
A:Gene: ptr

A:Map position: 61

C:Function:

A:Description: endopeptidase degrades small peptides [validated; MUID:92237263]  
A:Pathway: protein degradation  
C:Superfamily: insulysin  
C:Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein de-  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-962/Product: pitriylsin #status experimental <MAT>  
F:88,92/Binding site: zinc (His) #status experimental  
F:91/Active site: Glu #status experimental

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
Best Local Similarity 26.3%; Pred. No. 12;  
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;  
QY 126 WSRNEDITEPQSILAAAEKAGMSAEQA-----QGULEKIATPKVKNQ 167  
Db 731 WCRNKDVVDKKQSVIFEKAGNSTDSALAAVFVPTGYDEYTSAYSSLLGOIVQWPFYQ 790  
QY 168 LKETTEAACRYGAGLPITVAHVGDQTHMLFGSDRMELLALHGEKMGPIPPA 221  
Db 791 LR--TEBQLGVAVFAPFMSVGRQWGMGFLQSDNQ---PSFLWERYKAFFPTA 839

RESULT 11

I49343

probable isomerase doxJ - Pseudomonas sp. (strain C18)

C:Species: Pseudomonas sp.

C>Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 18-Sep-1998  
C:Accession: I49343; S27640

R:Denome, S.A.; Stanley, D.C.; Olson, E.S.; Young, K.D.

J. Bacteriol. 175, 6890-6901, 1993

A:Title: Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains:  
A:Reference number: A49343; MUID:94042852

A:Status: preliminary

A:Accession: I49343

A:Molecule type: DNA

A:Residues: 1-199 <DEN>

A:Cross-references: GB:M60405

C:Genetics:

A:Gene: doxJ

C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 7.5%; Score 89; DB 2; Length 199;

Best Local Similarity 24.2%; Pred. No. 1.7;

Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

QY 8 VELFYDVLSPYSWLGFEILCRYQNIWNINLQLRPSLITGIMKDSGNKPPGL--LPRKGLY 65

Db 3 VDFYDFLSPFSLYANHLRLSKLAQDYGFIRYAYDLARVKIAGNVGSPNRDLIVKLDY 62

QY 66 MANDLKLRLHQLIPIHPKDFLSVMLEKGSLL--SAMRELTAVNLEHPEMLEKASRELW 122

Db 63 LKVDLQWAELEYETPLVEPANYNRRMTNLTLYSGAMAQTGAYVNV-----VF 110

QY 123 MWVWSRNEDITEPQSILAA--AEKAGMSAEQGLEKATPKVKNOLKETTEAACRYGA 180

Db 111 NAVW--GDGIAPLESPLVSEKLGWDRSAFE---DFISSDAATERYDEQTHAAIERKV 165

QY 181 FGLPITVAHVGDQTHMLFGSDRMELLALHLLG 211

Db 166 FGVP-TMFLGD---EMWGNDRLEFLENVAVG 192

RESULT 12

T51712

threonine dehydratase/deaminase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000

C:Accession: T51712

R:Mourad, G.S.; Emerick, R.M.; Marion, A.L.; Smith, A.M.

Query Match 7.5%; Score 89; DB 2; Length 592;  
Best Local Similarity 22.8%; Pred. No. 6.9;  
Matches 49; Conservative 29; Mismatches 57; Indels

[illegible]

RESULT 15  
B71301  
probable tRNA delta(2)-isopentenylpyrophosphate transferase (miaA) - syphilis sp1  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
C:Accession: B71301  
R:Praser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.  
rson, J.; Khatak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A:title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770  
A:Accession: B71301  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

```

C:Superfamily: delta(2)-isopentenylpyrophosphate transferase

Query Match      7.1%;   Score 84.5;   DB 2;   Length 316;
Best Local Similarity 23.3%;   Pred. No. 7.4;
Matches 41;   Conservative 30;   Mismatches 60;   Indels 45;   Gaps 8;

QY 13 DVLSPYSLGFEILCRQYNTWINLQLRPSLLITGMKDSGNKRPGLLPRKGLYMANDLKL 72
   || || : : : : : || : : : : :
Db 80 DVCDPYE--EYANVFREQAAVYGI-----VPSIL-----RAHKVPPIIVGGTGLYLD---AV 124
   || || : : : : : || : : : : :
QY 73 LRHHLOIPIHFFPKDFLSVLMLEKGSLSAM-----RFLTAVNL- 108

```

[illegible]

Tue Apr 3 11:51:25 2001

us-09-441-723-1.rpr

Page 6

Search completed: April 3, 2001, 11:39:25  
Job time: 65 sec

---

Query Match	73.38;	Score 869;	DB 1;	Length 225;
-------------	--------	------------	-------	-------------





DR PROSITE; PS00143; INSULINASE; 1.  
KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal.  
FT SIGNAL 1 23  
FT CHAIN 24 962 PROTEASE III.  
FT METAL 88 88 ZINC.  
FT ACT\_SITE 91 91  
FT METAL 92 92 ZINC.  
FT METAL 169 169 ZINC.  
FT MUTAGEN 88 88 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 91 91 H->Q: LOSS OF ACTIVITY.  
FT MUTAGEN 92 92 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 162 162 E->Q: 20% LOSS OF ACTIVITY.  
FT MUTAGEN 169 169 E->Q: LOSS OF ACTIVITY AND OF ZN-BINDING.  
FT MUTAGEN 204 204 E->Q: NO LOSS OF ACTIVITY.  
FT CONFLICT 277 284 IIHVVPA -> HVHSLRPW (IN REF. 4).  
SQ SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
Best Local Similarity 26.3%; Pred. No. 4.4;  
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;

QY 126 WSRNEDITEPQSILAAAEKAGNSAEQA-----OGLLEKIATPKVKNQ 167  
Db 731 WCRNKVVVDKQSVIFEKAGNSDLSALAAVFTGYDEYTSAYSSLLGQVQPMFYNQ 790  
QY 168 LKETEACRYAGFLPITVAHVGDQTHMLFGSDRMELLAHLGKWMGPIPPA 221  
Db 791 LR--TEEQLYAVAFPMVSGRWGNGFLQSNQKQ---PSFLWERYKAFFPTA 839

RESULT 6  
NAHD\_PSESP STANDARD; PRT; 212 AA.  
ID NAHD\_PSESP  
AC Q52462;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 20-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE (HCCA ISOMERASE).  
GN DOXJ.  
OS Pseudomonas sp. (strain C18).  
OG Plasmid.  
OC Bacteria; Proteobacteria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94042852; PubMed=8226631;  
RA Denome S.A., Stanley D.C., Olson E.S., Young K.D.;  
RT "Metabolism of dibenzothiophene and naphthalene in Pseudomonas  
strains: complete DNA sequence of an upper naphthalene catabolic  
pathway.";  
RL J. Bacteriol. 175:6890-6901(1993).  
CC -!- FUNCTION: CATALYZES THE ISOMERIZATION OF 2-HYDROXYCHROMENE-2-  
CARBOXYLATE (HCCA) TO TRANS-O-HYDROXYBENZYLIDENEPYRUVATE (THBPA).  
CC THE OPTIMUM PH FOR THE ENZYME IS 10. THE REACTION IS REVERSIBLE.  
CC -!- PATHWAY: UPPER NAPHTHALENE CATABOLIC PATHWAY WHICH INVOLVES  
CC CONVERSION OF NAPHTHALENE TO SALICYLATE, AND CATABOLISM OF  
CC DIBENZOTHIOPHENE (DBT) AND PHENANTHRENE. CONVERTS PHENANTHRENE TO  
CC 1-HYDROXY-2-NAPHTHOIC ACID AND THE METABOLISM OF DBT IS LIMITED TO  
CC OXIDATION OF THE AROMATIC RING.  
CC -!- MISCELLANEOUS: DOXH AND DOXJ ENCODE DIFFERENT ENZYMES THAT MAY  
CC HAVE INTERCHANGABLE FUNCTIONS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M60405; AAA16133.1; -;  
KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
SQ SEQUENCE 212 AA; 24039 MW; 5EF96A619913DB4F CRC64;

Query Match 7.5%; Score 89; DB 1; Length 212;  
Best Local Similarity 24.2%; Pred. No. 0.75;  
Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

QY 8 VELFYDVLSPYSWLGFEILCRYQNIWNINQLRPSLITGIMKDSGNKPPGL--LPRKGLY 65  
Db 16 VDFYDFLSPYSYLANHRLSKLAQDYGFSIRYVYDAIDLARVKIAGNVCSPNRDLIVKLDY 75  
QY 66 MANDKLLRHHLQIPIHFPKDFLSVNLKGLS--SAMREFLTA-VNLEHPEMLEKASRELW 122  
Db 76 LKVDLQRAELYEIPLVFPANYSRRMTGLYSGAMAQTCAYVNV-----VF 123  
QY 123 MRYWSRNEDETEPQSILAA--AEKAGMSAEQAQGLLEKIATPKVKYNQLKETEACRYCA 180  
Db 124 NAVW--GDGTADLESLPALVSEKLGWDRSAFE---DFISSDAATERVDEOTHAATERKV 178  
QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLG 211  
Db 179 FGYP-TWFLGD--EMWGNDRLEFMLENAV 205

RESULT 7  
SG2\_BOVIN STANDARD; PRT; 613 AA.  
ID SG2\_BOVIN  
AC P20616;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SECRETOGRAININ II PRECURSOR (SGII) (CHROMOGRANIN C).  
GN SCG2 OR CHGC.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90264409; PubMed=2345170;  
RA Fischer-Colbrie R., Gutierrez J., Hsu C.M., Iacangelo A., Elden L.E.;  
RT "Sequence analysis, tissue distribution and regulation by cell  
depolarization, and second messengers of bovine secretogranin II  
(chromogranin C) mRNA.";  
RL J. Biol. Chem. 265:9208-9213(1990).  
CC -!- FUNCTION: SECRETOGRAININ II IS A NEUROENDOCRINE SECRETORY GRANULE  
CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.  
CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY  
CC GRANULES.  
CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
CC -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRAININ PROTEIN  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; J05468; AAA30760.1; -;  
DR PIR; A35296; A35296.  
DR INTERPRO; IPR001990; -;  
DR PFAM; PF01271; Grainin; 1.  
DR PROSITE; PS00422; GRANINS\_1; 1.  
KW Sulfatation; Cleavage on pair of basic residues; Calcium-binding;  
KW SIGNAL.  
FT SIGNAL 1 27 POTENTIAL.  
FT PROPEP 28 30 POTENTIAL.  
FT CHAIN 31 613 SECRETOGRAININ II.  
FT PEPTIDE 181 213 UNKNOWN ACTIVITY PEPTIDE (PROBABLE).  
FT MOD\_RES 150 150 SULFATATION (BY SIMILARITY).  
SQ SEQUENCE 613 AA; 70356 MW; 5DC079F559D83516 CRC64;





RX MEDLINE=91043032; PubMed=2146510;  
 RA Walker R.A., Salmon E.D., Endow S.A.;  
 RT "The Drosophila claret segregation protein is a minus-end directed  
 RL motor molecule.";  
 RN Nature 347:780-782(1990).  
 [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94155838; PubMed=8112290;  
 RA Lockhart A., Cross R.A.;  
 RT "Origins of reversed directionality in the ncd molecular motor.";  
 RL EMBO J. 13:751-757(1994).  
 [6]  
 RP MUTANT ALLELE NCD(D).  
 RX MEDLINE=91122049; PubMed=1825056;  
 RA Komma D.J., Horne A.S., Endow S.A.;  
 RT "Separation of meiotic and mitotic effects of claret  
 RL non-disjunctional on chromosome segregation in Drosophila.";  
 RL EMBO J. 10:419-424(1991).  
 [7]  
 RP CHARACTERIZATION OF MUTANT ALLELE NCD(D).  
 RX MEDLINE=96283629; PubMed=8670831;  
 RA Moore J.D., Song H., Endow S.A.;  
 RT "A point mutation in the microtubule binding region of the Ncd motor  
 RL protein reduces motor velocity.";  
 RL EMBO J. 15:3306-3314(1996).  
 [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 335-700.  
 RX MEDLINE=96195067; PubMed=8606780;  
 RA Sablin E.P., Kull F.J., Cooke R., Vale R.D., Fletcher R.J.;  
 RT "Crystal structure of the motor domain of the kinesin-related motor  
 RL ncd.";  
 RL Nature 380:555-559(1996).  
 CC -1- FUNCTION: NCD IS REQUIRED FOR NORMAL CHROMOSOMAL SEGREGATION IN  
 CC MEIOSIS, IN FEMALES, AND IN EARLY MITOTIC DIVISIONS OF THE EMBRYO.  
 CC THE NCD MOTOR ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S MINUS  
 CC END.  
 CC -1- MISCELLANEOUS: NCD(D) IS A MUTANT ALLELE THAT SHOWS ABNORMAL  
 CC CHROMOSOMAL SEGREGATION.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X52814; CAA36998.1; -;  
 DR EMBL; M33932; AAA28716.1; -;  
 DR EMBL; AE003771; AAF56942.1; -;  
 DR EMBL; X57475; CAA40713.1; -;  
 DR PIR; A35624; A35624.  
 DR PIR; S09748; S09748.  
 DR HSP; P17119; 3KAR.  
 DR FLYBASE; FBgn0002924; ncd.  
 DR INTERPRO; IPR001752; -;  
 DR PFAM; PF00225; kinesin; 1;  
 DR PRINTS; PRO0380; KINESINHEAVY.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;  
 -KW Meiosis; Mitosis.  
 FT DOMAIN 196 346 COILED COIL (POTENTIAL).  
 FT DOMAIN 347 700 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).  
 FT NP\_BIND 434 441 ATP (BY SIMILARITY).  
 FT MUTAGEN 556 556 V->F: IN NCD(D); REDUCES MOTOR VELOCITY.  
 FT CONFLICT 697 697 S -> N (IN REF. 1).  
 SQ SEQUENCE 700 AA; 77473 MW; ADE043CECFD561 CRC64;

Query Match

7.0%; Score 83.5; DB 1; Length 700;

Best Local Similarity 22.4%; Pred. No. 9.6;  
 Matches 52; Conservative 37; Mismatches 70; Indels 73; Gaps 12;  
 QY 1 MGPLPRTVELFVDVLSYSLWG--FELCRYQNTWNINLQLRPSLIITGIMKD-----SG 52  
 DB 451 VGVIPRTVDLLFDSIRYENLNGWEYEIKATFLEIYN---EVLVDLLSNEQKDEIRMAKN 507  
 QY 53 NKPPGLLPKRGLYMAN---DLKLLRHHLQIPIHPKDFLSVMLEKGSLSAMRFLTAVNLE 109  
 DB 508 NK-----NDIYVSNITEETVLPDNPHLRLHMTAK-----MNRATAS-----TAGN-- 547  
 QY 110 HPEMLEKASRELMMRWMSRNEDETEPOSILAAAKAGMS-----AEOAGLLEK 158  
 DB 548 -----ERS-----SRSHAVTKLEIGRHAEKQETSVGSINLVDLAGSESPTSTRM 593  
 QY 159 IATPKVKNLKETTEACRYGAGPLPTVAHVVDGQTHMLFGSDRMELLAHL 210  
 DB 594 TETNINRSLSELTN-----VILALLQKQDHIPY---RNSKLTHLL 631  
 RESULT 12  
 SG2\_RAT STANDARD; PRT; 619 AA.  
 ID SG2\_RAT  
 AC P10362;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SECRETAGRANIN II PRECURSOR (SGII) (CHROMOGRANIN C).  
 GN SG2 OR CHGC OR SCG-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89098327; PubMed=3211750;  
 RA Gardes H.-H., Phillips E., Huttner W.B.;  
 RT "The primary structure of rat secretogranin II deduced from a cDNA  
 RL sequence.";  
 RL Nucleic Acids Res. 16:11811-11811(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93309708; PubMed=8321414;  
 RA Kakar S.S., Wei N., Mulchahey J.J., Leboeuf R.D., Neill J.D.;  
 RT "Regulation of expression of secretogranin II mRNA in female rat  
 RL pituitary and hypothalamus.";  
 RL Neuroendocrinology 57:422-431(1993).  
 CC -1- FUNCTION: SECRETAGRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE  
 CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.  
 CC -1- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY  
 CC GRANULES.  
 CC -1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
 CC -1- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETAGRANIN PROTEIN  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X13618; CAA31950.1; -;  
 DR EMBL; M93669; AAA42135.1; -;  
 DR PIR; S02180; S02180.  
 DR INTERPRO; IPR001990; -;  
 DR PFAM; PF01271; Granin; 1.  
 DR PROSITE; PS00422; GRANINS\_1; 1.  
 KW Sulfatation; Cleavage on pair of basic residues; Calcium-binding;  
 KW Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 619  
 FT PEPTIDE 184 216 SECRETAGRANIN II.  
 FT UNKNOWN ACTIVITY PEPTIDE (PROBABLE).

FT MOD\_RES 153 153 Sulfatation (by similarity).  
SQ SEQUENCE 619 AA; 71031 MW; 27CB75B4F25A38D1 CRC64;

Query Match 6.9%; Score 81.5; DB 1; Length 619;

Best Local Similarity 24.8%; Pred. No. 13; Mismatches 57; Indels 51; Gaps 8;  
Matches 41; Conservative 16;

QY 13 DVLSPYSWLGPEILCRYQNIWNINLQRPISLITGIMKDSGNKPPGGLPRKGLYMANDLKL 72

DB 105 DVLSEDEW-----RIILEALQAEENEPSPALKENKPYALMLEK- 143

QY 73 LRHHLQIPHFPPKDFLSVLMKXGSLSAMRFLTAVNLEHPMLEKASRELMRVWSNEDI 132

DB 144 -----NFPVDPDDVETQWPERKLMRE-----PLMYEENSRE--NPKRTNEI 187

QY 133 TE-----POSILAAAEKAGMSAQGLLEKIATPKVKNQKLETTE 173

DB 188 VEEQYTPQS-LATLESVFOE-----LGKLTGP--SNQKRERVD 222

RESULT 13

DPOL\_ADE12

ID DPOL\_ADE12 STANDARD; PRT; 1061 AA.

AC P06538;

DT 01-JUN-1988 (Rel. 06, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE DNA POLYMERASE (EC.2.7.7.7).

GN POL.

OS Human adenovirus type 12.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

RN [1]

RX MEDLINE=94076430; PubMed=8254750;

RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;

RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative

functional analysis."

RL J. Virol. 68:379-389(1994).

RN [2]

RX MEDLINE=87106854; PubMed=3803925;

RA Shu L., Hong J.S., Wei Y.-F., Engler J.A.;

RT "Nucleotide sequence of the genes encoded in early region 2b of human

adenovirus type 12."

RL Gene 46:187-195(1986).

CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -

CC -1- N PYROPHOSPHATE + DNA(N).

CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.

CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X73487; CAA51882.1;

DR EMBL; M14785; AAA42478.1; ALT\_INIT.

DR PIR; A25770; DJAD12.

DR PIR; S33933; S33933.

DR INTERPRO; IPR002064; -.

DR PRINTS; PR00106; DNPOLB.

DR PROSITE; PS00116; DNA\_POLYMERASE\_B.1.

KW Transferase; DNA-directed DNA polymerase; DNA replication;

KN DNA-binding.

FT CONFLICT 32 32 R -> S (IN REF. 2).

FT CONFLICT 162 162 V -> L (IN REF. 2).

FT CONFLICT 181 182 LQ -> YN (IN REF. 2).

FT CONFLICT 461 461 S -> T (IN REF. 2).

FT CONFLICT 575 575 L -> F (IN REF. 2).

FT

FT CONFLICT 892 892 S -> T (IN REF. 2).  
FT CONFLICT 1030 1030 K -> M (IN REF. 2).  
SQ SEQUENCE 1061 AA; 121727 MW; 33FBA89C33065C08 CRC64;

Query Match 6.9%; Score 81.5; DB 1; Length 1061;

Best Local Similarity 21.8%; Pred. No. 25;

Matches 50; Conservative 35; Mismatches 77; Indels 67; Gaps 12;

QY 1 MGPLPRTVELF--YDVLSPYSWLGPEILCRYQNIWNINLQRPISLITGIMKDSGNKPPGL 58

DB 133 IGSHPRTFLRFDYDV-ETVTWVG-----AFGKQLVPPFML--VMKLSG----- 172

QY 59 LPRKGLYMANDLKLRRHLLQIPI-----HFPKD---FLSVMLEKXGSLSAMRFLTAVNLEH 110

DB 173 -----DDNLVKHALQALALELGWDQWERDSTTFYCLTPEKMKV-GQFRTYRN--- 218

QY 111 REMLEKASRELMRVWSNEDITEPOSILAAAEKAGMSAQGLLEKIATPKVKNQKLE 170

DB 219 -RLQTSLATDLWMTFLQKNPHLSQ-----WAOENGIVALEDLISYEDLKRAPIKGEPR- 271

QY 171 TTEACRYGAFGLPITVAHVQDQTHMLFSGDRMELLAHLHLLGKWMGP 219

DB 272 -----FVELYIVG--HNINGFDEIVLAAQVINNRDVP 304

RESULT 14

YG73\_SYNV3

ID YG73\_SYNV3 STANDARD; PRT; 274 AA.

AC P74261;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HYPOTHETICAL TRNA/RRNA METHYLTRANSFERASE SLR1673 (EC 2.1.1.-).

GN SLR1673.

OS Synchocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

Synchocystis sp. strain PCC6803. II. Sequence determination of the

entire genome and assignment of potential protein-coding regions."

RL DNA Res. 3:109-136(1996).

CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRMH FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; D90913; BAA18355.1; -.

DR INTERPRO; IPR001537; -.

DR PFAM; PF00588; Spo\_methylase; 1.

KW Hypothetical protein; Transferase; Methyltransferase.

SQ SEQUENCE 274 AA; 29701 MW; A4176C1061CAAE88 CRC64;

Query Match 6.8%; Score 81; DB 1; Length 274;

Best Local Similarity 21.1%; Pred. No. 5.1;

Matches 35; Conservative 32; Mismatches 61; Indels 38; Gaps 6;

QY 7 TVELFYDVLSPYSWLGPEILCRYQNIWNINLQRPISLITGIMKDSGNKPPGLPRKGLVM 66

DB 107 TMDQFWRSPPPHARLGL-VLERLQDPGNLGTILRTAATGV-----EGIWL 151

Search completed: April 3, 2001, 11:49:17  
Job time: 632 sec

```
Query Match      6.8%; Score 80.5; DB 1; Length 339;
Best Local Similarity 21.7%; Pred. No. 7.3;
Matches 65; Conservative 25; Mismatches 96; Indels 113; Gaps 14;
```

OY	3	PLPRTVEFDVLS---	PYSWLGFEILCRYQNIWNINLQRPSLIITCMKDSGNKPGLL	59
		: :	: :	
Ddb	47	LPTSPMNLKCQSAPPTS	WLKRFLFCRPPHWK-----SLRTASXSSESTKAALK	99
OY	60	PRK-GLYMANDULKLRHHH	LOIPIHFPKDFLSVMLEKGSL-----SAMRFUTAVNLEH	110
		:	:	
Ddb	100	PRPGSGWMTRKRSSPRS	SALGRAWH--PDRRGLVDGKGQVRSLADUETQACNAFAAIN-KA	157
		:	:	
OY	111	PMLE---KASRELMMRVMS	RNEIDTEPISQILAAASKAG-----	146
		: :	: :	
Ddb	158	PAILEGEVEFEREV----	SVIAARDSGNVAFIDLAENVHKDGLATST	202
OY	147	-----NSAEQAQGLEKIA-	-----TPKVKNOLKETTE	173
		:    :	:    :	
Ddb	203	VPAATSQVTAEATAEA	KALLHALDYGVGLUEFFVLKDGTLLANEFAPVHNS-GHWTE	261
OY	174	AACRYGAF-----GLPI--	-TVAHVDQTHMLFGSD-----RMELLAHLGECK	213



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compuqen Ltd.

OM protein - protein search, using sw.model

Run on: April 3, 2001, 11:38:20 ; Search time 25.94 Seconds  
(without alignments)  
1021.164 Million cell up

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPIPTVEIFYDVLSPYSW.....AHLIGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%  
Maximum Match 100%

Maximum Match 100%  
Listing first 45 summaries

**Database :**

```

SPTREMBL15:*
1:  sp_archea:*
2:  sp_bacteria*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_inverteb*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle*
9:  sp_phase:*
10: sp_plant:*
11: sp_robot:*
12: sp_virus:*
13: sp_vertebrn*
14: sp_unclasse*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	1185	100.0	226	4	Q9Y2Q3	Q9Y2Q3 homo sapien
2	1169	98.6	226	4	Q9P1S4	Q9P1S4 homo sapien
3	164.5	13.9	194	2	Q52782	Q52782 rhizobium l
4	149	12.6	195	2	Q9X3Q7	Q9X3Q7 sphingomona
5	145.5	12.3	312	5	Q22312	Q22312 caenorhabdi
6	140	11.8	197	2	Q85994	Q85994 sphingomona
7	135	11.4	136	2	Q9ZHH4	Q9ZHH4 burkholderi
8	114	9.6	196	2	Q9WXH0	Q9WXH0 alcaligenes
9	106	8.9	199	2	Q9Z3X5	Q9Z3X5 pseudomonas
10	97.5	8.2	238	2	Q9Z167	Q9Z167 pseudomonas
11	91	7.7	581	2	Q9WXQ0	Q9WXQ0 thermotoga
12	90.5	7.6	410	1	Q26322	Q26322 methanobact
13	89	7.5	199	2	Q51499	Q51499 pseudomonas
14	89	7.5	592	10	Q9SPF1	Q9SPF1 arabidopsis
15	89	7.5	592	10	Q9ZSS6	Q9ZSS6 arabidopsis
16	83.5	7.0	308	2	Q9ZNM7	Q9ZNM7 exiglobacte
17	83.5	7.0	1007	10	Q9ZVD4	Q9ZVD4 arabidopsis
18	83	7.0	863	12	Q93124	Q93124 human calic
19	82.5	7.0	263	3	Q87197	Q87197 thermus aqu

## ALIGNMENTS

RESULT	1
Q9Y2Q3	
PRELIMINARY;	PRT; 226 AA.
Q9Y2Q3;	
OAC	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE	GLUTATHIONE S-TRANSFERASE SUBUNIT 13 HOMOLOG.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Mao M., Ye M., Zhang Q., Zhou J., Wu J., Shen Y., Kan L., He K.,
RA	Gu B., Fu G., Chen S.; Chen Z.;
RA	"Human RGSTK1-1 homolog gene.";
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF070657; AAD20963.1; -
KW	Transference.
SQ	SEQUENCE 226 AA; 25497 MW; D3FDAFDI533B58A4 CRC64;
Query Match	100.0%; Score 1185; DB 4; Length 226;
Best Local Similarity	100.0%; Pred. No. 4.9e-96;
Matches 226; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MGPLPRTVLEFYDVLSPYSWLGFEILCRYQNIWININQLRPSLTITGIMKDSGNKPGLLP 60       
Ddb	1 MGPLPRTVLEFYDVLSPYSWLGFEILCRYQNIWININQLRPSLTITGIMKDSGNKPGLLP 60       
QY	61 RKGLYMNDLKLLRRHLQTPIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120       
Ddb	61 RKGLYMNDLKLLRRHLQTPIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRE 120       
QY	121 LWMRWVSNRNEDITEPQSIIAAAEKAGMSAQOGLLEKIATPKVNQLKETTEAACRYGA 180       
Ddb	121 LWMRWVSNRNEDITEPQSIIAAAEKAGMSAQOGLLEKIATPKVNQLKETTEAACRYGA 180       
QY	181 FGLPTTAHVADGQTHMIFGSDRMELLAHLLEGKWMGPippAVNARL 226       
Ddb	181 FGLPTTAHVADGQTHMIFGSDRMELLAHLLEGKWMGPippAVNARL 226       

```

*
RESULT 2
Q9P1S4 ID Q9P1S4 PRELIMINARY; PRT; 226 AA.
AC Q9P1S4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HCDMD47P.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhao Z., Huang X., Li N., Zhu X., Cao X.;
RT "A novel gene from human dendritic cell.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF068287; AAF65506.1; -.
SQ SEQUENCE 226 AA; 25586 MW; D3FDB561533B5A65 CRC64;

Query Match 98.6%; Score 1169; DB 4; Length 226;
Best Local Similarity 99.1%; Pred. No. 1.2e-94;
Matches 224; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSPSYWLGFELCRQYQNIWININLQRLPSLITGIMKDSGNKPPGLP 60
Db 1 MGPLPRTVELFDVLSPSYWLGFELCRQYQNIWININLQRLPSLITGIMKDSGNKPPGLP 60
QY 61 RKGLYMANDLKLRRHLQIPIHPKDFLSVMEKGSLSAMRFLTAVNLEHPMLEKASRE 120
Db 61 RKGLYMANDLKLRRHLQIPIHPKDFLSVMEKGSLSAMRFLTAVNLEHPMLEKASRE 120
QY 121 LWMRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETEAACRYGA 180
Db 121 LWMRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETEAACRYGA 180
QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLGKWMGPPIPPAVNARL 226
Db 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLGKWMGPPIPPAVNARL 226

RESULT 3
Q52782 ID Q52782 PRELIMINARY; PRT; 194 AA.
AC Q52782;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE PUTATIVE GLYCEROL-3-PHOSPHATE TRANSPORT PROTEIN (UGPC).
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=384;
RN [1]
RP SEQUENCE FROM N.A.
RA Brito B., Palacios J., Imperial J., Ruiz-Argueso T.;
RL MEDLINE=96328256; PubMed=8765742;
RT "Identification of a gene for a chemoreceptor of the methyl-accepting
RT type in the symbiotic plasmid of Rhizobium leguminosarum bv. viciae
RT UPM791."
RL Blochim. Biophys. Acta 1308:7-11(1996).
DR EMBL; U23040; AAC4312.1; -.
SQ SEQUENCE 194 AA; 21153 MW; 0421C131B37BA435 CRC64;

Query Match 13.9%; Score 164.5; DB 2; Length 194;
Best Local Similarity 25.2%; Pred. No. 7.3e-07;
Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;

*
RESULT 4
Q9X9Q7 ID Q9X9Q7 PRELIMINARY; PRT; 195 AA.
AC Q9X9Q7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.
GN NSAD.
OS Sphingomonas sp.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Sphingomonas.
OX NCBI_TaxID=28214;
RN [1]
RP SEQUENCE FROM N.A.
RA Keck A.;
RC STRAIN=BN6;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U65001; AAD43416.1; -.
KW Isomerase.
SQ SEQUENCE 195 AA; 22064 MW; 7E6FBAC96B3C6316 CRC64;

Query Match 12.6%; Score 149; DB 2; Length 195;
Best Local Similarity 23.1%; Pred. No. 1.7e-05;
Matches 50; Conservative 43; Mismatches 91; Indels 32; Gaps 7;

QY 4 LPTVELFDVLSPSYWLGFELCRQYQNIWININLQRLPSLITGIMKDSGNKPPG---LLP 60
Db 1 MTKTIDYDFDFISPSYLAQVKLPDLARTGCVIERPIDIPEAKIAGNYGSPNREVVP 60
QY 61 RKGLYMANDLKLRRHLQIPIHPKDF-----LSVMLEKGSLSAMRFLTAVNLEHPML 114
Db 61 KIKVMA-DLERWAAKYEVPLTFPASFCSDMNCALYARGDQAFAVTA----- 111
QY 115 EKASRELWMRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETEA 174
Db 112 -----YHRIWGICIDPRDQNELRGCAEDVGLDAD---ALCEFVRSPPAGGEYRKARTQ 161
QY 175 ACRYGAGLPITVAHVGDQTHMLFGSDRMELLAHL 210
Db 162 AYQGVFGAP--MMFVDDQ--IFWGNDRDLDFLESYL 193

RESULT 5
Q22312 ID Q22312 PRELIMINARY; PRT; 312 AA.
AC Q22312;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE COSMID T07E3.
GN T07E3.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,  
RA Smalton S., Smith A., Sonhammer E., Staden K., Sulston J.,  
RA Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RL Nature 0:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Favello A.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U13643; AAA21082.1; -  
SQ SEQUENCE 312 AA; 35436 MW; 0FB8561F57687120 CRC64;

Query Match 12.3%; Score 145.5; DB 5; Length 312;  
Best Local Similarity 22.6%; Pred. No. 6.3e-05;  
Matches 51; Conservative 46; Mismatches 100; Indels 29; Gaps 6;  
QY 1 MGPLPRTVELFYDVLSPYSW-----LGFELCRYQ-----NIWNINLQ 38  
DB 1 MALPR-VKCYDFVCPNSWITQALTSNLSFERIDFEVDFKIGILHNAQWNRQ 59  
QY 39 LRPSLI--TGIMKDSGNKPPGLPRKGLYMAND---LKLRLHRLQIPIHPKDFLSVME 93  
DB 60 VHSRLWKTKIEVPEQAEEETLSEMGILQKIDERGKLGICERVVPVDPWKNYKTAVA 119  
QY 94 KGSLSAMRFLTAVNLEHPEMELEKASRELWMRVWSRNEDITEPOSILAAAEKAGNSAQ 153  
DB 120 RGSVIPQFITSIREQYDPLVEKATHHLGKRLWEQRLPVHYGCHMSTVCRELGISFKIAE 179  
QY 154 GLLEKIATPKVNOLKETTEAACRYGAFGLPITVAHVD--GQTHMLF 198  
DB 180 DIVARLSSPENRSILHNCKEAVDFKLEAPGLILLTDEGDTIKIF 225

RESULT 6  
O85994 PRELIMINARY; PRT; 197 AA.  
AC O85994;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.  
GN NAHD.  
OS Sphingomonas aromaticivorans.  
OG Plasmid pNLI.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
OC Sphingomonas.  
OX NCBI\_TaxID=48935;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FI99;  
RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,  
RA Sensen C.W., Gaasterland T., Saffer J.D., Fredrickson J.K.;  
RT \*Complete sequence of a 184 kb catabolic plasmid from Sphingomonas  
RT aromaticivorans strain FI99.\*  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF079317; AAD04010.1; -

KW Isomerase; Plasmid.  
SQ SEQUENCE 197 AA; 21998 MW; C0CD74B7106C07CE CRC64;  
Query Match 11.8%; Score 140; DB 2; Length 197;  
Best Local Similarity 23.1%; Pred. No. 0.0001;  
Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;  
QY 4 LPTVELFYDVLSPYSWLGFEELCRYQNIWNINLQRLPSLITGIMKDSGNKPPG---LLP 60  
DB 1 MTTIDIFYDFISFYSYLAQLKLPETIARAAGCTVDYWPIDPEAKIAAGNVPNSREVL 60  
QY 61 RKGLYMANDLKLRLHRLQIPIHPKDF-----LSVMLEKGSLSAMRFLTAVNLEHPEML 114  
DB 61 -KIKVMKADLERWAERYGVPLTFPASFACADWNCACVLFAREHGKAEAFV----- 109  
QY 115 EKASRELWMRVWSRNEDITEPOSILAAAEKAGNSAQAGLLEKIATPKVNOLKETTEA 174  
DB 110 -----DAYRRIWGOGIDPDRNELAACATAAGL---DPAALIAFVESPAGQNEYRKARSQ 161  
QY 175 ACRYGAFGLPITVAHVDGQTHMLFGSDRMELLAHL 210  
DB 162 AIQGVYGAQL--MFVDDQ--IFWGNDRDLDEFLAEYL 193  
RESULT 7  
Q9ZHH4 PRELIMINARY; PRT; 196 AA.  
AC Q9ZHH4;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE ISOMERASE PHND.  
GN PHND.  
OS Burkholderia sp. RP007.  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Burkholderia.  
OX NCBI\_TaxID=83784;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RP007;  
RX MEDLINE=99102216; PubMed=9882667;  
RA Laurie A.D., Lloyd-Jones G.;  
RT \*The phn genes of Burkholderia sp. strain RP007 constitute a divergent  
RL gene cluster for polycyclic aromatic hydrocarbon catabolism.\*;  
DR J. Bacteriol. 181:531-540(1999).  
KW EMBL; AF061751; AAD09871.1; -  
SQ SEQUENCE 196 AA; 21908 MW; 58716E9039BEA76B CRC64;

Query Match 11.4%; Score 135; DB 2; Length 196;  
Best Local Similarity 24.0%; Pred. No. 0.00028;  
Matches 50; Conservative 35; Mismatches 103; Indels 20; Gaps 5;  
QY 7 TVELFYDVLSPYSWLGFEELCRYQNIWNINLQRLPSLITGIMKDSGNKPPGL--LPRKGL 64  
DB 2 TIDFFDFLSPYAYLARHRLTQVAALHGCAIAYKPIDLAKAKLAIGNTPANRDMPVKLA 61  
QY 65 YMANDLKLRLHRLQIPIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMELEKASRELWMR 124  
DB 62 YVVEDLKRWAARYRIPIETIKNFNTKRMNVGT-----FYAARGQOQADYVROAYHLWGE 116  
QY 125 VWSRNEDITEPOSILAAAEKAGMS-AEQAGLLEKIATPKVNOLKETTEACRYGAFGL 193  
DB 117 GGAPDDD-----AALRSIAVSMGMDAADFLRFLDSSEAEATYNNESTLEAISAGVFCV 168  
QY 184 PITVAHVDGQTHMLFGSDRMELLAHL 211  
DB 169 PTMAVGRD----MMWGNDRIDFLETHLG 192  
RESULT 8

```

Q9WXH0
ID Q9WXH0 PRELIMINARY; PRT; 196 AA.
AC Q9WXH0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ISOMERASE.
GN PHND.
OS Alcaligenes faecalis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Alcaligenes.
OX NCBI_TaxID=511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AFK2;
RT Kiyohara H., Tabata Y., Takizawa N.;
RT "A phenanthrene degradative gene cluster in Alcaligenes faecalis
RL AFK2".
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB024945; BAW76325.1; -.
KW isomerase.
SQ SEQUENCE 196 AA; 22355 MW; C027579783BEE2E2 CRC64;

Query Match 9.6%; Score 114; DB 2; Length 196;
Best Local Similarity 21.1%; Pred. No. 0.019;
Matches 44; Conservative 40; Mismatches 99; Indels 26; Gaps 61

QY 8 VELFYDVLSPYSWLGFG----EILCRYQNIWINLQLRPSLITGIMKDSGNKPPGL--LPR 61
Db 2 LSFYDFVSPFSYLASIRLPEIVQRY---GISVSYKPIDIACAKRAIGNVGPSNRDMPV 57
QY 62 KGLYWANDLKLRLHHLQIPIHPPKDFLSVMLEKGSLSAMRPLTAVNLBHEPMLEKASREL 121
Db 58 KLTHLSRDLQWAQRYGTGTPKPPSPDSRLNTGFG----FYAAGEAREAEYVRR----- 108
QY 122 WMRYVSRNREDITEPOSILAAAEKAGSAEQAGLLEKIATPKVKYNOLKETTETAACRYGAF 181
Db 109 FHLTWGNGQASGEPVILRSIASEMGNVDDFMQFDTSV---DGANEYKQSIDEGIARSVF 165
QY 182 GLPITVAHVGDQTHLMFGSDRMELLAHL 210
Db 166 GVPWVVI-----GDEMWWGNDRDLDFVDEYL 190

```

```

RESULT      9
Q923X5
ID      Q923X5      PRELIMINARY;
AC      Q923X5;
DT      01-MAY-1999      (TrEMBLrel. 10, Created)
DT      01-MAY-1999      (TrEMBLrel. 10, Last sequence update)
DT      01-MAY-1999      (TrEMBLrel. 10, Last annotation update)
DE      2-HYDROXYCHROMENE CARBOXYLATE ISOMERASE.
GN      NAGD.
OS      Pseudomonas sp. U2.
OG      Plasmid pMWU2.
OC      Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC      Ralstonia.
OX      NCBI_TaxID=70356;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=U2;
RX      MEDLINE=98233751; PubMed=9573207;
RA      Fuemmayer S.L., Wild M., Boyes A.L., Williams P.A.;
RT      "A gene cluster encoding steps in conversion of naphthalene to
RT      gentisate in Pseudomonas sp. strain U2.";
RL      J. Bacteriol. 180:2522-2530(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=U2;
RA      Fuemmayer S., Wild M., Boyes A.L., Williams P.A.;
RL      Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF036940; AAD12617.1; -.

```

[illegible]

```

RESULT 10
Q92I67
ID Q92I67 PRELIMINARY; PRT; 238 AA.
AC Q92I67;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE DEHYDROGENASE.
GN NAHD.
OS Pseudomonas stutzeri (Pseudomonas perfectmarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI_TaxID=316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AN10;
RA Bosch R., Lalucat J., Timmis K.N., Moore E.R.B.;
RT "Complete nucleotide sequence of a chromosomally encoded naphthalene
RT degradation pathway from Pseudomonas stutzeri AN10 and its
RT evolutionary significance.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF039533; AAD0214.1; -.
SQ SEQUENCE 238 AA; 26359 MW; 29B9372C1CCFC71B CRC64;

```

```

Query Match      8.2%; Score 97.5; DB 2; Length 238;
Best Local Similarity 23.5%; Pred. No. 0.69;
Matches 53; Conservative 37; Mismatches 103; Indels 33; Gaps 9;

QY      8 VELEFDVLSPYSWLGFEEILCRQYNININLQLRPSLITGIMKDSCKNPGL--LPRKGLY 65
      | : : | | | | : | : : : : | | | | |
Db      19 VDFYDFDLPSPSYLANHRLSKLAHDYSFSIRYHSIDLARAKIATGNVGPSNRDLKVKLAY 78

QY      66 MANDLKLRLHHLQIPIHFPPKDFLSVMLEKG-SUSAMRFLTAVNLEHPEMLEKASRELWMR 124
      | : : | | | : | | : : | | | | |
Db      79 LMVDLKLWAEYGLPFLFPANTYSQRNMAGLYYSGAETQTAAVY-----NTVFNA 128

QY      125 WVSRNEDIT-LPQSTILAAEKAGMSAEQAQGLEKIATPKVKNQLKETTAAACRYCAGFL 183
      | | | : : | : | | : : | : : | | |
Db      129 VW--GEGIALDSESLALV--CGTLCWDRAAFEEFLSSDAATNAYDEHTQAAIERKVFGV 184

QY      184 PITVAHVDGQTHMFGSDRMELLALHL-----GEKWMGPI 218
      | : | : | : | : | : | : | : | : |
Db      185 P-TMFLGD---QMWNGNDRFLFLENTLRCCSGAGIAAAGETGVKPV 226

RESULT 11
99WXQ0

```

ID Q9WXQ0 PRELIMINARY; PRT; 581 AA.  
 AC Q9WXQ0;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
 DE ABC TRANSPORTER, ATP-BINDING PROTEIN.  
 GN TM0043.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RL Nature 398:323-329(1999).  
 DR EMBL: AE001691; AAD35137.1; -.  
 DR HSPF: P13569; INED.  
 DR TIGR: TM0043; -.  
 DR INTERPRO: IPR001140; -.  
 DR INTERPRO: IPR001617; -.  
 DR INTERPRO: IPR002106; -.  
 DR PFAM: PF00005; ABC\_tran; 1.  
 DR PFAM: PF00664; ABC\_membrane; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE: PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 KW ATP-binding.  
 SQ SEQUENCE 581 AA; 66324 MW; 55719E0DCAD5D0A05 CRC64;

Query Match 7.7%; Score 91; DB 2; Length 581;  
 Best Local Similarity 19.0%; Pred. NO. 8.4;  
 Matches 35; Conservative 44; Mismatches 67; Indels 38; Gaps 6;

QY 28 RYQNIW-----NINLQLRPSLTGIMKDSGNKPPGL-----LPRKGLYMAN 68  
 DB 343 RFENWFSYDGNWVKNVLDLQFQPKLYAIVGTEGGKSTLMSLINGLYTPQKNIFD 402  
 QY 69 DLKLRHLLQIP-----IHPPKDFL-----SVMLEKGSLSAMRFLTAVNLEHP-EMLE 115  
 DB 403 EIPLEYNLKLVRKQIAAQPQDVLLFSGTILNIRLEDESIPERVLEALKRVHLDIIE 462  
 QY 116 KASRELWVRVNSRNEDITEPQSILAAAEKAGM-----SAPQAGLLEKIAATPKVKNQIK 169  
 DB 463 RLPGGVYVEIVRGTTLSAGERQLIARAVLFDKIFILDEATSNVDVITETKIQEALE 522  
 QY 170 ETTE 173  
 DB 523 ELSK 526

RESULT 12  
 ID O26322 PRELIMINARY; PRT; 410 AA.  
 AC O26322;  
 DT 01-JAN-1998 (TReMBLrel. 05, Created)  
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
 DE FLAVOPROTEIN A HOMOLOG (II).  
 GN MTH220.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 OC Methanobacterium.  
 OX NCBI\_TaxID=2166;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=DELTA H;  
 RX MEDLINE=98037514; PubMed=9371463;  
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., DuBois J.,  
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL: AE000809; AAB84726.1; -.  
 DR INTERPRO: IPR001279; -.  
 DR PFAM: PF00753; lactamase\_B; 1.  
 SQ SEQUENCE 410 AA; 45740 MW; 8EAF9D3B363A8BD5 CRC64;

Query Match 7.6%; Score 90.5; DB 1; Length 410;  
 Best Local Similarity 25.1%; Pred. NO. 5.8;  
 Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;

QY 13 DVLSPYSWLGFEILC-----RYQNIWNINLQLRPSLTGIMKDSGNK-----S4  
 DB 99 EVLKRY---GSEICTAKAAEGLRQHSIPDTPMQ---TVKTGSDTLGGKTLTELEAP 152  
 QY 55 ----PPG---LLPRKGLYMANDLKLLRHLLQIPHPKDFLSVMLEKGSLSAMRF-----L 103  
 DB 153 MLHWPDSMPTLLEEGILFSND---AFQHLICISKREKDVPEAVLMD---AAMKFEYANLL 207  
 QY 104 TAVNLEHPMLEKAS--RELWM-----RVWSRNEDITEPQSILAAAEKAGMSA 149  
 DB 208 TPLS---PLVLRKSEVKELGKLEKIGMIAPSHQIWI-----TFPKLIIA-----YT 252  
 QY 150 EOAQGLLEKIAATPKVKNQIKETEAACRYGAFGLPITVAHVGDQTHMLFGSRMELLALH 209  
 DB 253 DWATCKCRDKAT-IYDTHMYSTRMLAHAMAEGE--MAADVDSMHFLHEDERSEIVKNI 309  
 QY 210 LGEK-----WMGPIP 219  
 DB 310 LESKAVFTGSPFMNGPFP 328

RESULT 13  
 ID Q51499 PRELIMINARY; PRT; 199 AA.  
 AC Q51499;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)  
 DE ISOMERASE.  
 GN PAHE.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PAK1;  
 RA Takizawa N., Iida T., Yamauchi K., Satoh S., Wang Y., Fukuda M.,  
 RA Kiyohara H.;  
 RT "The molecular analysis of an NAH7-type gene cluster, pah, located on  
 RT the chromosome of Pseudomonas aeruginosa PaK1.";  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: D84146; BAA12247.1; -.  
 KW Isomerase.  
 SQ SEQUENCE 199 AA; 22509 MW; 81C2A90CA56E5462 CRC64;

Query Match 7.5%; Score 89; DB 2; Length 199;  
 Best Local Similarity 24.2%; Pred. NO. 3;  
 Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

```

QY 8 VELFYDVLSPYSWLGFEILCRONTYIWINIQLRPSLITGIMKSGNKPGL--LPRKGLY 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 VDFYDFLSPYSYLANHRLSKLAQDYCFYSYAIIDLARVKIAICGNVGPSNRDLTVIKDY 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 MANDLKLRHHLLQIPIHPKDFLSVMLEKSL--SAMRFLTA-VNLEHPMLERASRELW 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 LKVDLQRWAEYIEIPLVFPANTNSRRMTGLIYSGAMAQTAGYVNV-----VF 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 MRVSRNEDITEPQSILAA--AEKAGMSAEQAQGLLEKIAITPKVNOLKETTEACRYGA 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 NAVV--GGIAPDLSELPAVSEKLCWDRSAFE--DFISSDAATERYDEQTHAAIERKV 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 FGLPITVAHVQGOQTHMFGSDRMELLAHLIG 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 FGVP--TMFLGD---EMWGNDRLFMLENAVG 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 14
ID Q9SPF1 PRELIMINARY; PRT; 592 AA.
AC Q9SPF1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE THREONINE DEHYDRATASE/DEAMINASE.
OS OMRI.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Rosidae;
OC Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GM11B;
RT Mourad G.S., Emerick R.M., Smith A.M.;
RT "Cloning and Sequencing of a cDNA Encoding an Isoleucine Feedback
RT Insensitive Threonine Dehydratase/Deaminase of Arabidopsis thaliana
RT line GM11b (omri/omri).";
RL Submitted (Aug-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF177212; AAC54324.1; -.
DR HSSP; P04968; 1TDJ.
DR INTERPRO; IPR000524; -.
DR INTERPRO; IPR000634; -.
DR INTERPRO; IPR001721; -.
DR INTERPRO; IPR001926; -.
DR PFAM; PF00291; PALP, 1.
DR PFAM; PF00585; Thr_dehydrat_C; 2.
DR PROSITE; PS00043; HTH_CNTR_FAMILY; UNKNOWN_1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
SO SEQUENCE 592 AA; 64562 MW; 1D02F73AD53FAF6C CRC64;

```

RESULT	15
Q92SS6	
ID	Q9ZSS6 PRELIMINARY; PRT; 592 AA.
AC	O9ZSS6;
DT	01-MAY-1999 (TrEMBLrel. 10, Created)
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE	THREONINE DEHYDRATASE/DEAMINASE [EC 4.2.1.16].
GN	OMRI OR T2K18.12.
OS	Arabidopsis thaliana (Mouse-ear cross).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC	Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA;
RA	Mourad G., Emerick R., Marion A., Smith A.;
RT	"Cloning and Sequencing of a cDNA Encoding Threonine
RT	Dehydratase/Deaminase of Arabidopsis thaliana (Accession No. AF096281)
RL	(PGR 98-199).";
RL	Plant Physiol. 118:1534-1534(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA;
RA	Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA	Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA	Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT	"Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";
RL	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	Mourad G.S., Smith A.M.;
RA	"Molecular characterization of the genomic clone, including the
RT	promoter sequences, of threonine dehydratase/deaminase from
RT	Arabidopsis thaliana.";
RL	Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
EMBL	EMBL; AF096281; AAC97936.1; -.
DR	EMBL; AC010927; AAF04418.1; -.
DR	EMBL; AF221984; AAF32370.1; -.
DR	HSP; P04968; 1TDJ
DR	INTERPRO; IPR000524; -.
DR	INTERPRO; IPR000634; -.
DR	INTERPRO; IPR001721; -.
DR	INTERPRO; IPR001926; -.
DR	Pfam; PF00291; PALP; 1.
DR	Pfam; PF00585; Thr_dehydrat_C_2.
DR	PROSITE; PS00043; HTH_GNTR_FAMILY; UNKNOWN_1.
DR	PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW	Lysate.
SQ	SEQUENCE 592 AA; 64634 MW; 16658747052FAE7C CRC64;

Search completed: April 3, 2001, 11:39:53  
Job time: 93 sec

---

